Neurocomputing 213 (2016) 125–136

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Fusion of multiple channel features for person re-identification

Xuekuan Wang [a,1], Cairong Zhao [a,b,c,1], Duoqian Miao [a,*], Zhihua Wei [a], Renxian Zhang [a], Tingfei Ye [c]

[a] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[b] Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information (NJUST), Ministry of Education, Nanjing 210094, China
[c] Key Laboratory of Cloud Computing and Intelligent Information Processing of Changzhou City, Jiangsu University of Technology, 213001 Changzhou, China

ARTICLE INFO

Article history:
Received 17 September 2015
Received in revised form
30 November 2015
Accepted 13 December 2015
Available online 29 July 2016

Keywords:
Person re-identification
Multiple channel
Feature fusion

ABSTRACT

Person re-identification plays an important role for automatic search of a person's presence in a surveillance video, and feature representation is a critical and fundamental problem for person re-identification. Besides, an reliable feature representation should effectively adapt to the changes of illumination, pose, viewpoint, etc. In this paper, we propose an effective feature representation called fusion of multiple channel features (FMCF) which captures different low-level features from multiple channels of HSV color space, considering the characteristics of different color channels and fusing color, texture and correlation of spatial structure. Furthermore, it takes advantage of an overlapping strategy to eliminate contrast of local cells in an image. In addition, we apply the simple weight distance metric to measure the similarity of different images, rather than metric learning which relies on a specific feature and requires more computing resources. Finally, we apply the proposed method of FMCF on the i-LIDS Multiple-Camera Tracking Scenario(MCTS) and CUHK-01person re-identification datasets, and the experimental results demonstrate that it is more robust to the variation of visual appearance.

1. Introduction

The task of person re-identification (re-id) is to match pedestrian images observed from disjoint views in non-overlapping camera networks [1]. It has attracted more and more attention in recent years due to its important applications in video surveillance. The key issue of re-id system is to capture reliable and robust features from pedestrian images and measure the similarity among them to estimate if they are from the same person. However, the complexity of the environment, which is affected by illumination, pose, viewpoint, occlusion, image resolution and camera setting in non-overlapping camera system, leads to various challenges [2]. At present, the state-of-the-art approaches for person re-identification are mainly divided into two groups: (1) the appearance-based approach designing of distinctive and stable descriptors to represent the person's appearance; (2) the metric learning approach obtaining a suitable metric method which minimizes the distance of the same person and maximizes the distance of different persons.

Most existing appearance-based approaches concern low-level features such as color (color histogram, Dominant color, color space, etc. [3–5,45]), texture (local binary pattern (LBP), Gabor, Co-occurrence matrix [6-8,35,40], etc. and shape [9–11,44,47]. These features are always combined to improve the recognition rate. Aiming to seek a distinctive and stable feature expression, researchers have proposed a lot of feature representation algorithms, ranging widely from symmetry-driven accumulation [12], covariance descriptor [13], horizontal stripe-based partition [14], pyramid matching [15], graph matching [16,48], salience matching [10,17], local maximal occurrence [18], hash model [41], sparse learning model [46,49,50], deep learning model [19–21], etc. Meanwhile, feature extraction and multi-feature fusion are two main issues for feature representation. In [22], D. Gray and H. Tao propose the method of ensemble of local features (EFL), achieving an efficient and intelligent descriptor for viewpoint invariant pedestrian recognition. While, in [23], the authors design the approach of bag of ensemble colors to combine low-level color histogram and semantic color names to represent human appearances. These handcrafted or learning-based descriptors have made impressive improvements over robust feature representation, and advanced the person re-identification research. Unfortunately, it is also extremely difficult to extract a stable feature which effectively adapts to serve changes and misalignment across disjoint views.

Another aspect of person re-identification is how to learn a robust distance or similarity function to deal with the complex matching problem. Many metric methods simply choose a standard distance such as l_1–norm [24], l_2 – norm – based distance

* Corresponding author.
E-mail addresses: zhaocairong@tongji.edu.cn (C. Zhao), dqmiao@tongji.edu.cn (D. Miao).

[1] The authors contribute equally to this work.

http://dx.doi.org/10.1016/j.neucom.2015.12.140
0925-2312/

[25], or Bhattacharyya distance [22]. However, they would essentially treat all features equally without learning and discarding bad features selectively, thus the matching result is always undesirable. In contrast, the distance learning based approaches typically learn a discriminative metric between appearance features of the same person and different persons across camera pairs. These methods mainly include Rank SVM [26], least squares [43], Relative Distance Comparison (RDC) [14], Kernel-Based Metric [27], Mahalanobis distance learning [29], Deep Metric Learning (DML) [30], metric ensembles [28], non-convex model [42], and Iterative re-weight sparse ranking [31] etc. In practice, many previous metric learning methods show a two-stage procedure for metric learning, that is, the Principle Component Analysis (PCA) is first applied for dimension reduction, then metric learning is performed on the PCA subspace. However, this two-stage procedure may not be optimal for metric learning in a low-dimensional space, because samples with different classes may already be cluttered after the first stage.

In this paper, we propose an effective appearance-based feature representation called fusion of multiple histogram features (FMCF). Different from the previous works, FMCF pays more attention to the properties with different channels of color space *HSV* and captures the weight-color histogram, texture and spatial structural information using color space component instead of gray-value images. Main contributions in the current paper are as follows:

A. A new feature descriptor called fusion of multiple channel features (FMCF) has been proposed.
B. The proposed method of FMCF captures color information form hue and saturation components of HSV color space, while texture and spatial structural information are extracted from value component.
C. Joint histogram, acquisition and matching have been done for person re-identification.
D. Histogram is captured with overlapping strategy from three components of HSV color space.

The remainder of this paper is organized as follows. We review the related works in Section 2. We introduce the theory of the proposed approach in Section 3. In Section 4, we carry out the comparative experiments on two public person re-identification datasets and give the detailed discussion based on the experimental results. Finally, conclusions are offered in Section 5.

2. Related work

This paper aims to seek an effective appearance-based method for person re-identification from the view of multiple channel features extraction. Therefore, first of all, we present an overview of the relevant works, i.e., HSV color space, local binary pattern (LBP) [33] and histogram of Oriented Gradients (HOG) [34].

2.1. HSV color space

In general, there are three types of images, including binary image, gray scale image and color image. Aiming to reduce the complexity of algorithm, most approaches just concern the gray scale image. In contrast, the color image consists of multiple channels which contain a range of intensity and describe more rich information. Among them, the RGB color space composed of red, green and blue is widely utilized for image representation. However, the three channels pay more attention to the color property, ignoring other characteristics of color space. Hence, in our proposed approach, we consider the HSV color space that stands for hue, saturation and value which describe different characteristics of color image.

Hue component is directly related to the color which can be distinguished by the human eye. It is defined as an angle value, varying from 0° to 360°. Each number corresponds to a different color. Saturation component describes the purity of color component, and the value shows the intensity of a color. It is numbered from 0 to 1, as it goes from low to high intensity of color. Meanwhile, value component also varies from 0 to 1 and it is most similar to gray-scale image.

Compared with other color spaces, many research works have explained that HSV color space is more effective for extracting color, intensity and brightness from images [38]. Besides, different channels of HSV color space are relatively independent and have weak correlation, which is conducive to extracting more varied and sufficient information from single color channel. Meanwhile, it can ensure that the dimension of feature vector is low. In our proposed approach, images are converted from the RGB space to the HSV color space, besides we deal with the characteristics of different channels to capture suitable and abundant features which combine the color, texture and spatial structural information for person re-identification.

2.2. Local binary pattern (LBP)

A landmark representative of these structural image descriptors is Local Binary Pattern first proposed by Ojala et al. [33] as a gray-scale invariant texture descriptor. LBP code is obtained by thresholding its circularly symmetric n-neighbors in a circle of radius r with a pixel value of central point, and arranging the results as a binary string, shown in Fig. 1. It is stable and robust for the change of illumination. The mathematical representation of *LBP* is as follows:

$$LBP_{n,r} = \sum_{i=1}^{n} P_1(I_i - I_c) \times 2^{i-1} \tag{1}$$

$$P_1(t) = \begin{cases} 1, & t \geq 1 \\ 0, & else \end{cases} \tag{2}$$

$LBP_{n,r}$ obtains the local binary pattern of each pixel, where n and r are the number of neighboring pixels and the radius of circle, taken for computation. I_i is the pixel value of center point and I_c is the pixel value of neighboring pixels.

2.3. Histogram of Oriented Gradients (HOG)

The Histogram of Oriented Gradients (HOG) is a popular descriptor that was initially proposed for pedestrian detection by Dalal and Triggs [34]. HOG is represented by the 3D histogram of gradient locations and orientations, and employs both rectangular and log-polar location grids. The process of generating HOG Descriptor for an image is shown in Fig. 2. Aiming to capture the HOG feature, the gradient and director of the images are compute by Eqs. (3)–(6).

$$G_x(x, y) = I(x+1, y) - I(x-1, y) \tag{3}$$

$$G_y(x, y) = I(x, y+1) - I(x, y-1) \tag{4}$$

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{5}$$

$$\alpha(x, y) = \tan^{-1}(G_x(x, y)/G_y(x, y)) \tag{6}$$

There, $I(x, y)$, $G_x(x, y)$, $G_y(x, y)$, $G(x, y)$ and $\alpha(x, y)$ represent color pixel value, the gradient of horizontal director, the gradient of vertical director, gradient values and directions at point (x, y),

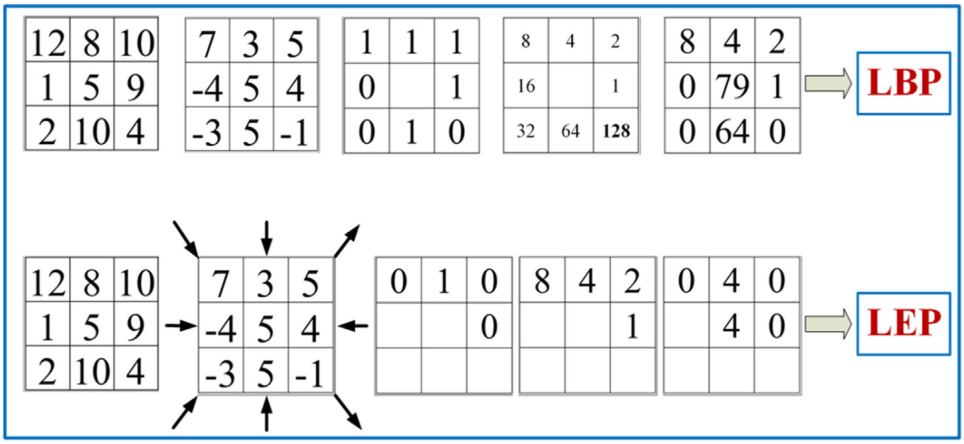

Fig. 1. Local binary pattern and local extrema pattern examples.

respectively.

Then, the authors divide the image windows into small spatial regions called cell and accumulate a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell. Aiming to better adopt the change of illumination, shadowing, contrast-normalization is applied to these local histograms. Then, these cells are grouped together in 2×2 overlapping blocks and each of these blocks is normalized individually. The overlapping of blocks ensures that there is no loss of local variations and consistency across the whole image. Finally, the HOG feature consists of these histograms captured from blocks.

3. FMCF: a novel image feature extraction method

As discussed in Section 1, the previous works, such as HOG and LBP, capture reliable features from gray-value images and take into account that the three channels of HSV color space are identical, ignoring the characteristics of different channels. While HSV color space consists of hue, saturation and value parts which separately represent different properties of a color image. We can extract more information from different channels, keeping low dimension of the feature vector. Inspired by this idea, we concern the characteristics of different channels of HSV color space and design suitable features to capture rich information, joining them by weight metric distance for person re-identification. It is more robust to the change of illumination and view, because it combines the advantages of HOG and LBP. The process of our proposed method is shown in Fig. 3 and we will introduce it in detail.

3.1. S-weight histogram of H-channel (SWH-H)

In the *HSV* color space, hue component corresponds to the color information and the value of hue is defined as an angle, varying from 0° to 360°. It is suitable to extract color information. In our proposed approach, we quantify the value of hue component into 12 bins, aiming to capture optimum color information and reduce the dimension of feature vector. In addition, saturation component represents the purity of color component and the value describes the intensity of a color. Based on this idea, we extract the histogram of color information from *H*-channel and apply the value of saturation as the weight of hue's value instead of equality. It is defined as:

$$h(H(x, y)) = \sum\sum h(H(x, y)) + S(x, y) \tag{7}$$

Where $H(x, y)$ is the value of H-channel with quantifying at point $I(x, y)$, and $S(x, y)$ is the value of S-channel. We can obtain S-weight histogram of H-channel (SWH-H) for description of the color information.

3.2. Local dominant orientation with V-channel (LDO-V)

As mentioned above, the value component of HSV color space is most similar to gray scale image which contains rich texture information, hence we take advantage of the local dominant orientation to extract the reliable gradient orientation

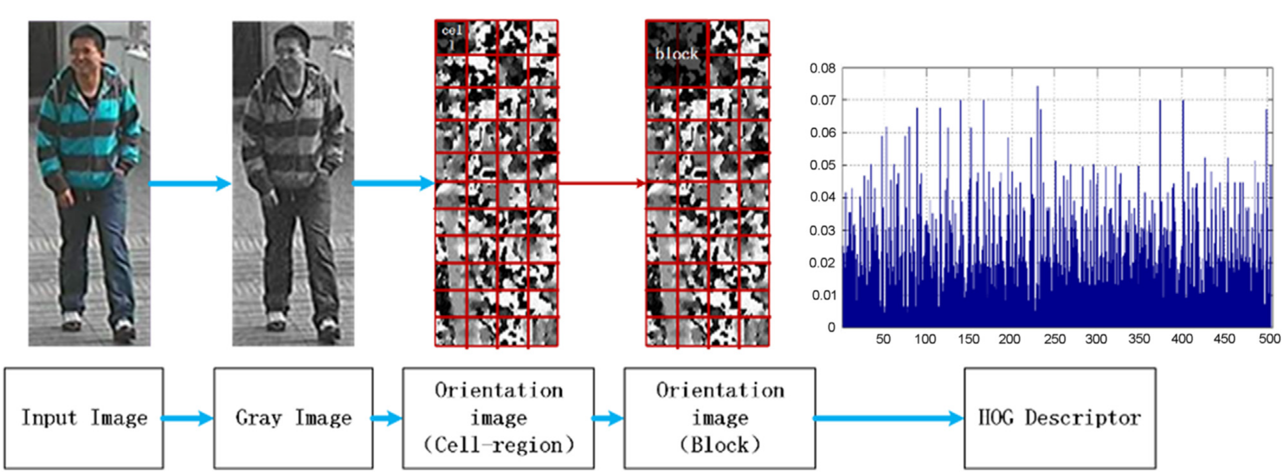

Fig. 2. Process of formation of Histogram of Oriented Gradients feature descriptor.

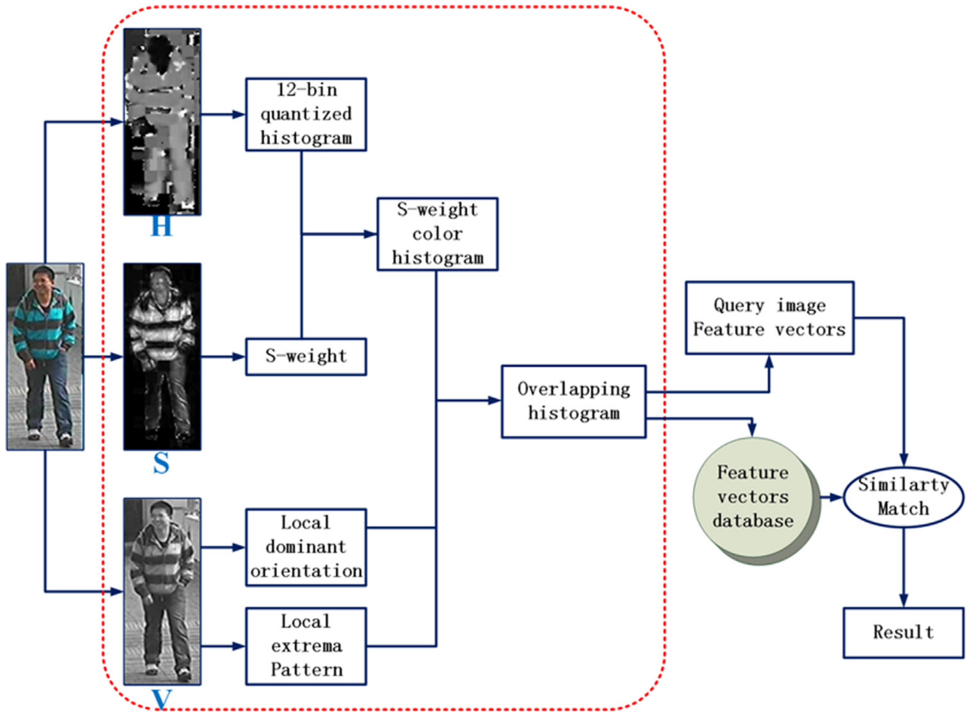

Fig. 3. An overview of the proposed method.

representation based on principal component analysis (PCA) method. PCA provides a set of optimal basis vectors to represent the given data and results in the minimum mean-square approximation error. It can be done by eigenvalue decomposition of the data covariance matrix or singular value decomposition (SVD) of data matrix [37]. Here, we apply the method of SVD to capture local dominant orientation.

We denote a gradient matrix over a $p \times p$ window (w_i) around the interesting point $I(x, y)$ of an image:

$$G = \begin{bmatrix} \dots & \dots \\ g_x(k) & g_y(k) \\ \dots & \dots \end{bmatrix}, k \in w_i \tag{8}$$

where $g_x(k)$ and $g_y(k)$ are the gradients of the image at point (x_k, y_k) in x and y directions, respectively. Then, we extract the reliable local information of the path w_i from the gradient matrix G. It is obtained by performing *SVD* on G.

$$G = USV^T = Udiag[s_1, s_2][v_1, v_2]^T \tag{9}$$

where U is a $P \times 2$ matrix, and V is a 2×2 matrix. For each matrix, the column vectors are orthogonal. S is a diagonal 2×2 singular value matrix representing the energy in the dominant orientation and its perpendicular direction. In the matrix of V, the first column vector $v_1 = [u_{1,1}, u_{1,2}]$ describes the dominant orientation of the local gradient field. Hence, the dominant orientation angle θ is defined as follows:

$$\theta = \arctan(v_{1,1}/v_{1,2}) \tag{10}$$

Furthermore, the singular values s_1, s_2 represent the energy information, the relative energy e of the dominant orientation in the path w_i is defined as follows:

$$e_i = \frac{s_1 + \lambda}{s_2 + \lambda} \tag{11}$$

where λ is the regularization parameter, which is used to keep the denominator of the ratio from being close to zero and retrain the effect of noise.

For every pixels $V(x, y)$ in V-channel, we will apply the local dominant orientation operator (*LDO-V*) to compute the dominant orientation and energy $(\theta_{x,y}, e_{x,y})$ and obtain the dominant orientation and energy map $O = \begin{bmatrix} (\theta_{1,1}, e_{1,1}) & \dots & (\theta_{1,y}, e_{1,y}) \\ \dots & \dots & \dots \\ (\theta_{x,1}, e_{x,1}) & \dots & (\theta_{x,y}, e_{x,y}) \end{bmatrix}$.

3.3. Local extrema patterns

The dominant orientation and its corresponding relative energy reveal the structural information in a local path. However, the operator pays more attention to principal information of a local path and ignores the correlation of different points in the local path. Hence, we take advantage of the local extrema pattern (LEP) which is designed by Murala et al. [36] and inspired by local binary pattern (LBP) operator to capture more texture and orientation information. LBP code considers each pixel (except boundary pixels) as a center pixel and is computed by a threshold of every neighborhood pixel with each center pixel. Differently, LEP deals with edge information in different directions, including 0°, 45°, 90°, 135°, and assigns 1 if both neighboring pixels are greater or less separately as compared to the center pixel in a particular direction, and 0 if otherwise , shown in Fig. 1. For a center pixel I_c and the corresponding neighbor pixel I_i, LEP is obtained as follows:

$$I_i' = I_i - I_c, i = 1, 2, \dots, 8 \tag{12}$$

$$I_i'(\psi) = P_2(I_k', I_{k+4}'), k = (1 + \psi/45), \psi \in \{0^o, 45^o, 90^o, 135^o\} \tag{13}$$

$$P_2(I_k', I_{k+4}') = \begin{cases} 1, & I_k' \times I_{k+4}' \ge 0 \\ 0, & \text{else} \end{cases} \tag{14}$$

$$\text{LEP}(I_c) = \sum_{\psi} 2^{\psi/45} \times I_k'(\psi) \tag{15}$$

In contrast, the operator of LEP can particularly describe the spatial correlation of a center point and its neighborhood points. Therefore, we apply the operator of LEP to extract texture, orientation and spatial structural information with the V-channel of

HSV color space, and make up for the loss information of the *LDO-V* operator.

3.4. Extracting overlapping histogram

It is well known that a single global histogram ignores the whole structural and spatial information of the object. Therefore, it cannot meet the demand of image feature extraction. Fortunately, the local information is effectively described by the local texture and color histogram. Besides, the LEP coding retains the stable local spatial structure which is appropriate to represent the appearance of pedestrians that has obvious structural characteristics. Inspired by the framework of HOG, we divide the image into a series of $k \times k$ cell-regions $\begin{bmatrix} c_{1,1} & \cdots & c_{1,n} \\ \cdots & \cdots & \cdots \\ c_{m,1} & \cdots & c_{m,n} \end{bmatrix}$ to obtain sufficient histogram features and make full use of overlapping strategy to construct the block $B_{x,y} = \left[c_{x,y}, c_{x+1,y}, c_{x,y+1}, c_{x+1,y+1} \right]$ for extracting the whole structural information of the appearance of pedestrian. Where k is the size of cell-regions and n, m represent the number of cell-regions in the horizontal and vertical directions, respectively. And $B_{x,y}$ represents the block which consists of 2×2 cell-regions. Then, each of these blocks is normalized individually and we capture the statistical information from these blocks.

In our proposed approach, we capture three 1-D local histograms from a $k \times k$ cell-region, including SWH-H, LDO-V and LEP-V, and respectively construct three different groups of features, shown in Fig. 4, rather than fuse them together. We denote these threes feature vectors as follows:

$$H_{swh-h} = \left[h_{swh-h}^{(1)}, h_{swh-h}^{(2)}, \dots, h_{swh-h}^{(L)} \right] \tag{16}$$

$$H_{ldo-v} = \left[h_{ldo-v}^{(1)}, h_{ldo-v}^{(2)}, \dots, h_{ldo-v}^{(L)} \right] \tag{17}$$

$$H_{lep} = \left[h_{lep}^{(1)}, h_{lep}^{(2)}, \dots, h_{lep}^{(L)} \right] \tag{18}$$

3.5. Ensembles of distance

In our proposed approach, we apply Canberra distance [38] to calculate the similarity between two feature vectors T and Q. It is shown in Eqs. (19)–(21):

$$D(T, Q) = \sum_{i=1}^{M} \frac{|T_i - Q_i|}{|T_i + u_t| + |Q_i + u_q|} \tag{19}$$

$$u_t = \sum_{i=1}^{M} T_i/M \tag{20}$$

$$u_q = \sum_{i=1}^{M} Q_i/M \tag{21}$$

Where M is the dimension of feature vectors. For three feature vectors of H_{swh-h}, H_{ldo-v} and H_{lep}, we can obtain three different distances, denoted as D_{swh-h}, D_{ldo-v} and D_{lep}. Finally, a weight distance metric is conducted via:

$$D = \alpha \times D_{swh-h} + \beta \times D_{ldo-v} + \gamma \times D_{lep} \tag{22}$$

Where weighted coefficients α, β and γ are set as 0.5, 0.3, 0.2, respectively.

3.6. The overall algorithm

Algorithm: Fusion of multiple channel features

Input: a gallery set $G = \{g_i\}$, $i = 1, 2, \dots, n$ and a probe set $P = \{p_j\}$, $j = 1, 2, \dots, m$.

Begin:

(1) Convert RGB color space to HSV color space, obtaining G' and P';
(2) Quantify the value of H-channel into 12-bins;
(3) Extract S-weight color histograms with H-channel (SWH-H) from cells;
(4) Extract local dominant orientation histograms with V-channel (LDO-V) from cells;

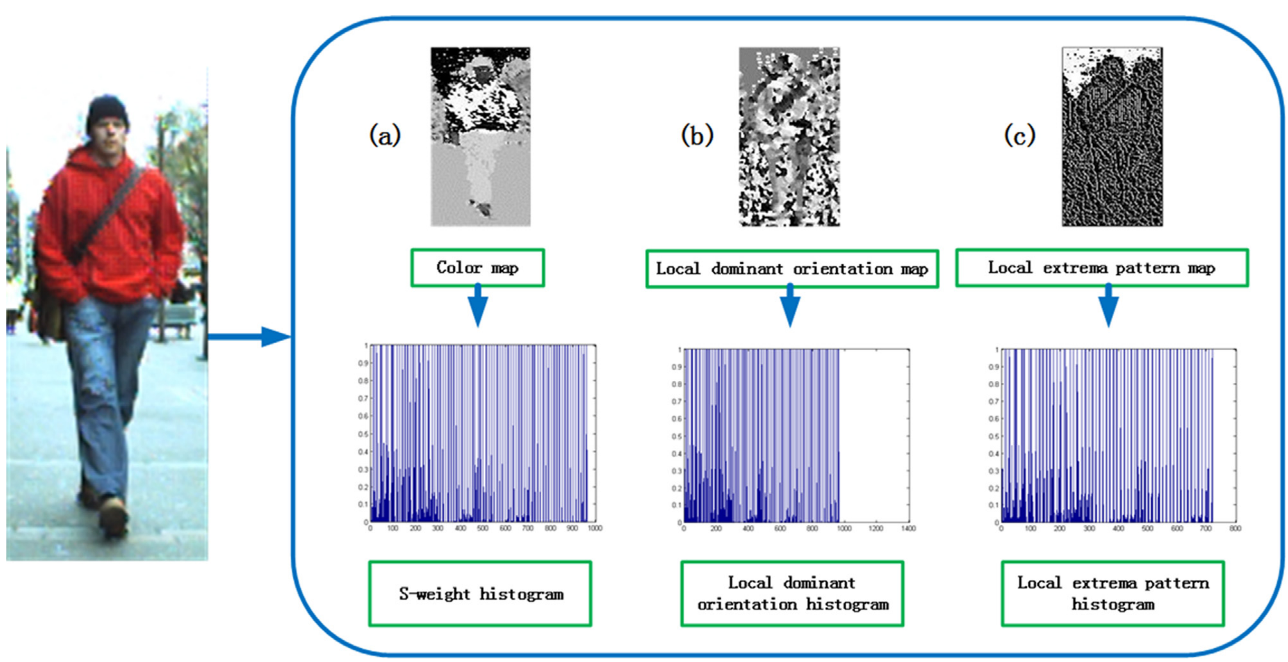

Fig. 4. The histograms captured from different channels considering the characteristics of HSV color space: S-weight histogram extracted from H-channel and S-channel, local dominant orientation histogram extracted from V-channel, and local extrema pattern histogram extracted from V-channel.

Algorithm: Fusion of multiple channel features

(5) Extract local extrema pattern histograms with V-channel (LEP-V) from cells;
(6) Construct blocks and obtain histograms with SWH-H, LDO-V and LEP-V, respectively, denoted as H_{swh-h}, H_{ldo-v},H_{lep-v};
(7) Obtain the set of feature vectors with gallery and probe, denoted as H^G_{swh-h}, H^G_{ldo-v}, H^G_{lep-v} and H^P_{swh-h}, H^P_{ldo-v}, H^P_{lep-v};
(8) Obtain the Canberra distance between gallery and probe, denoted as $D^{p,g}_{swh-h}$, $D^{p,g}_{ldo-v}$ and $D^{p,g}_{lep-v}$;
(9) Obtain the finally distance: $D_{p,g} = \alpha \times D^{p,g}_{swh-h} + \beta \times D^{p,g}_{ldo-v} + \gamma \times D^{p,g}_{lep-v}$;
(10) Obtain the matching rate.

End

Output: the matching rate.

4. Experiments

In this section, we compare the performance of the proposed approach (FMCF) to that of the state-of-the-art methods reported on i-LIDS Multiple-Camera Tracking Scenario (MCTS) and CUHK-01 person re-identification datasets. The i-LIDS MCTS dataset is used to testify the performance over variance of lighting illumination, obstacle, etc. The CUHK-01 dataset concerns the problem of large-view angle change and illumination change can also be drastic. For assessment of performance, we randomly choose all images of p persons (classes) to set up the test set which includes a gallery set and a probe set. The gallery set consists of one image form each person and the remaining images are used as the probe set. This procedure is repeated 10 times.

Meanwhile, we use the standard performance measurements to evaluate our proposed approach (FMCF), also known as Cumulative Matching Characteristic (CMC) and Synthetic Disambiguation/Re-acquisition Rate (SD/RR) curves [32]. The CMC curve represents the expectation of the probe image correct match at rank r against the p gallery images. And rank-1 matching rate is thus the correct matching, recognition rate. However, the SD/RR curve measures the probability that any of the m best matches is correct. In practice, a high rank-1 matching rate is significant, meanwhile, the top r ranked matching rate with a small r value is also critical because the top matching images will normally be verified by a human operator. Next, we will report and analyze the detailed results of experiment with the problem of person re-identification.

4.1. Performances on i-LIDS MCTS dataset

In i-LIDS MCTS dataset, which is captured at an airport arrival hall in the busy times in a multi-camera CCTV network, there are a total of 119 persons and 476 images, shown in Fig. 5. All images are normalized to 90×160 pixels in our experiments. Besides, a lot of these images undergo the change of view angle and large illumination, and are subject to large occlusions. Next, we will analyze our proposed approach (FMCF) in details with this dataset.

4.1.1. Comparison with the relative algorithms

In this experiment, we choose all images containing $p = 30, 50, 80$ persons to test the performance of the proposed approach (FMCF), compared with the relative methods of color histogram, LBP [33], HOG [34], ELF [22], LEP [36] and PRICoLBP [39]. The matching rate shown in Fig. 6(a1, a2, a3) gives an idea of the trend of the curve across all ranks. Meanwhile, in Table 1, we can see that our method achieves a rank-1 matching rate of 42.7%, 37.7%, and 31.7% with $p = 30, 50, 80$, outperforming the best result obtained by ELF, which achieves a rank-1 matching rate of 33.4%, 30.0%, and 25.0%. It is shown that our proposed approach captures color, texture and spatial structural information, and integrates the advantages of HOG and LEP to reduce intra-class variations. Besides, the overlapping strategy enhances the robustness of illumination variation, so that the same person can be recognized at a higher rank. Furthermore, from the SD/RR curves shown in Fig. 6 (b1, b2, b3), we can see that the performance of our proposed approach (FMCF) is also superior to that of others. This indicates that any of the m-best matches is correct than the other approaches on this dataset.

Besides, different scales of cell for capturing local 1-D histogram will lead to extracting different features which have different performances for person re-identification in this dataset. So, what is the suitable sizes in practice? Next, we will further analyze the effects on performance, undergoing different scales of cell.

4.1.2. Experiments using different sizes of cell

In this section, we apply different sizes of cell in our proposed method (FMCF) to extract 1-D histogram feature and carry out experiments. These lead to different results, the performances at rank-1, 5, 10, and 20 are listed in Table 2. Note that, the size of cell

Fig. 5. Examples of person re-identification on i-LIDS MCTS.

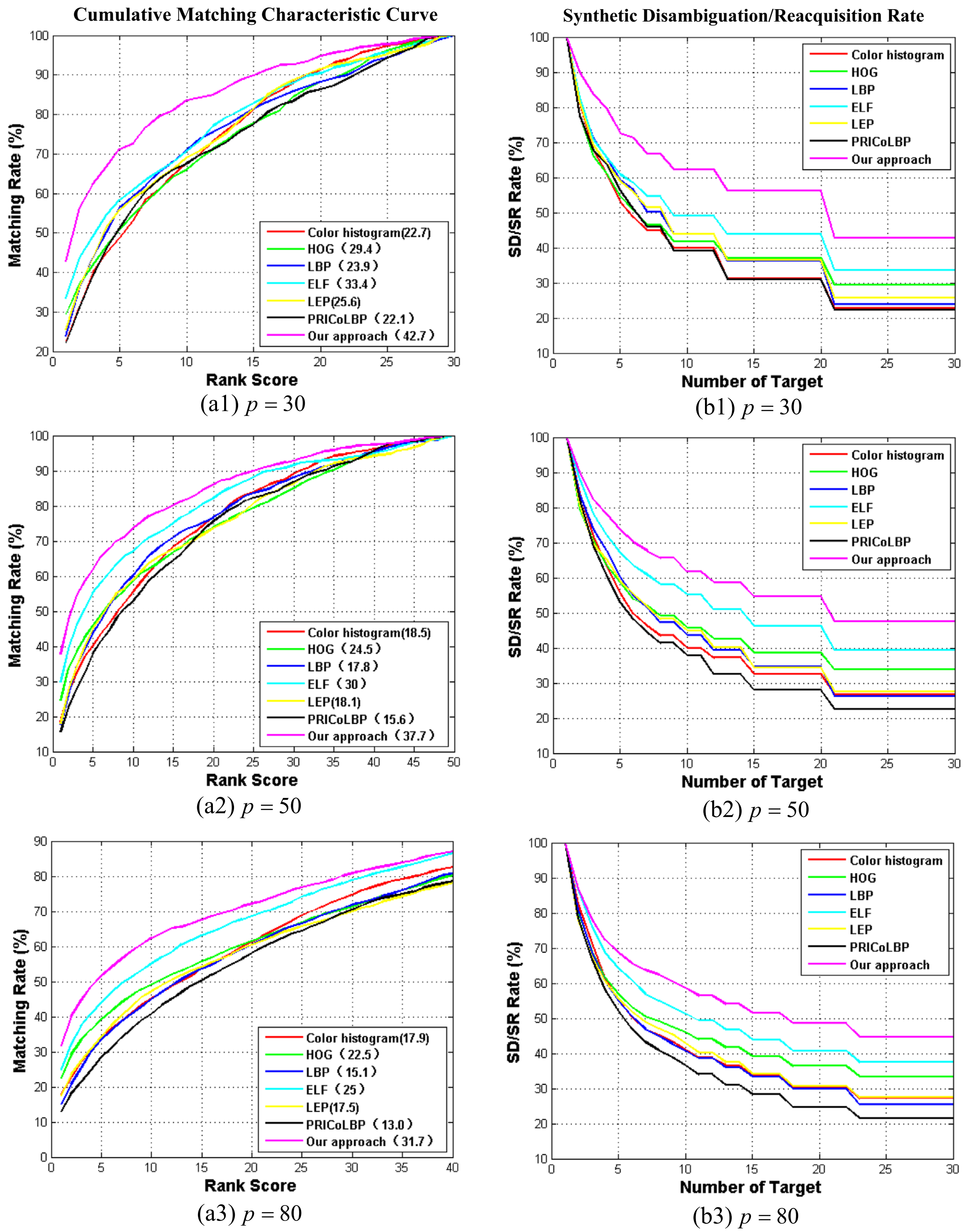

Fig. 6. The CMC and SD/RR curves of our approach, color histogram, LBP, HOG, ELF, LEP and PRICoLBP on i-LIDS MCTS dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as 15×16 probably outperforms other cells, although the matching rate at rank-20 is not the best. In order to eliminate the local variations and ensure a lower dimension of feature vector which can retain more information, we empirically take advantage of the size of cell as 15×16 in our experiments.

4.1.3. Analysis of the fusion of multiple feature

Our proposed approach (FMCF) concerns the characteristics of different channels with HSV color space and makes full use of most suitable descriptors to describe the features of multiple channels. Aimed to intuitively show the improvement of the performance, we report the rank 1, 5, 10, 20 matching rate (%) with our approach (SWH-H + LDO-V + LEP-V), SWH-H, LDO-V, LEP, SWH-H + LDO-V on i-LIDS MCTS dataset, shown in Fig. 7 and Table 3. We can see that the descriptor of SWH-H captures histogram features from S-channel and H-channel and the matching rate is higher than the descriptors of LDO-V and LEP-V which all extract texture and structural information from V-channel.

Table 1
The rank 1, 5, 10, 20 matching rate (%) with our approach, color histogram, LBP, HOG, ELF, LEP and PRICoLBP on i-LIDS MCTS dataset.

Methods	$p=30$				$p=50$				$p=80$			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
Our	42.7	71.3	83.6	94.9	37.7	61.7	73.6	86.0	31.7	51.5	62.5	72.2
Color histogram	22.7	48.6	67.7	91.3	18.5	40	55.4	76.7	17.9	33.5	45.3	61.4
HOG	29.4	50.7	66	88.3	24.5	45.8	58.6	73.7	22.5	39.1	49.1	61.6
LBP	23.9	56.4	71	88.3	17.8	43.6	60.1	75.6	15.1	33.3	45.0	60.6
ELF	33.4	58.3	70.7	90.4	30	55.3	67.0	82.2	25	43.9	55.1	68.7
LEP	25.6	55.7	69.1	91.6	18.1	44.9	59.0	73.6	17.5	34.2	47.2	60.6
PRICoLBP	22.1	51.0	67.6	86.4	15.6	37.9	52.9	77.2	13.0	28.4	40.8	58.1

Table 2
The rank 1, 5, 10, 20 matching rate (%) with different sizes of cell oni-LIDS MCTS dataset.

Person		$p=30$														
Cell	**Width**	15					30					45				
	Height	16	20	32	40	80	16	20	32	40	80	16	20	32	40	80
Rate (%)	**Rank-1**	**42.7**	38.6	40.4	35.6	29.7	**40.7**	**39**	**37.4**	34.3	27.9	38.6	38.9	35.9	32.4	29.1
	Rank-5	**71.3**	69.1	69	69	64.7	70.1	**70.6**	67.3	68.4	65.6	**69.7**	68.9	67.9	66.7	61.9
	Rank-10	**83.6**	79.7	80	79.4	76.7	81.6	82.4	**77.4**	80.6	78.9	**80.3**	80.4	81.9	79.7	77.3
	Rank-20	**94.9**	93	92.4	93.9	89.4	94.1	**94.1**	**95.3**	95	91.6	92.4	90.9	92.3	90.6	91.4

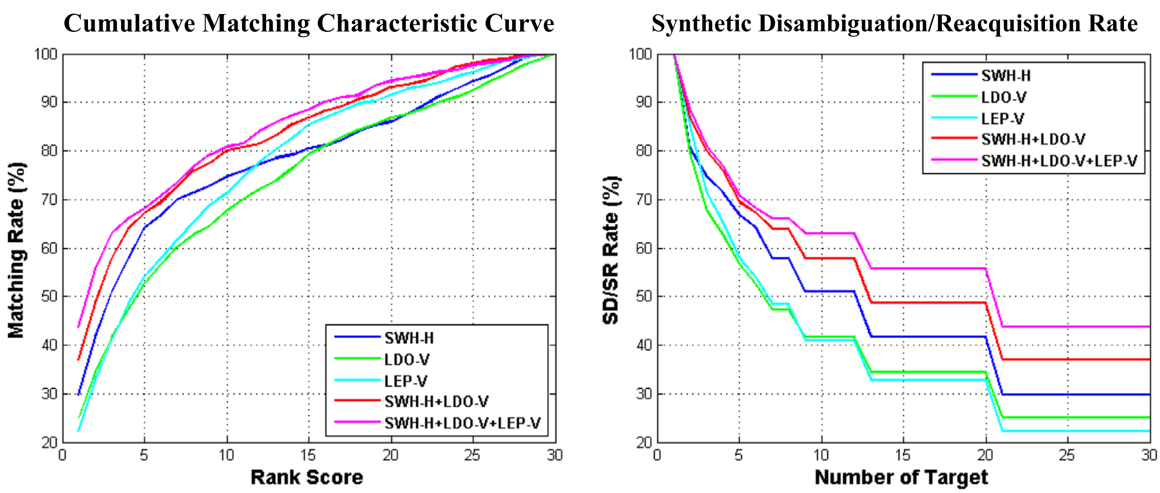

Fig. 7. The CMC and SD/RR curves of our approach (SWH-H + LDO-V + LEP-V), SWH-H, LDO-V, LEP-V and SWH-H + LDO-V on i-LIDS MCTS dataset.

Table 3
The matching rate (%) with our approach (SWH-H + LDO-V + LEP-V), SWH-H, LDO-V, LEP-V and SWH-H + LDO-V on i-LIDS MCTS dataset.

Methods	$p=30$							
	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	$r=10$	$r=15$	$r=20$
SWH-H	29.7	41.7	51	57.7	64.1	74.7	80.6	86
LDO-V	25	34.3	41.6	47.1	52.6	67.7	79.4	86.9
LEP-V	22.3	32.9	40.9	48.4	54	71.3	85.4	91.6
SWH-H +LDO-V	37	48.7	57.9	63.9	67.1	80	87	93.1
SWH-H +LDO-V +LEP-V	**42.7**	**56.3**	**62.3**	**66.7**	**71.3**	**83.6**	**90**	**94.9**

Combing the descriptors of SWH-H and LDO-V, it can capture more color and texture information of HSV color space, so that the performance is superior to a single descriptor, achieving the matching rate of 37% at rank=1. While the descriptor of LDO-V only extracts the local dominant orientation information, ignoring the correlation of spatial structure in the local cells. Hence, our proposed approach fuses the descriptors of SWH-H, LDO-V and LEP-V to capture color, texture and spatial correlation from different channels, and it is more robust to the changes of view angle and large illumination, etc. Compared with the descriptor of SWH-H + LDO-V, it has a more effective performance for person re-identification on this dataset, achieving the rank-1 matching rate of 42.7%.

4.1.4. Analysis of the parameters of weight distance metric

Our proposed approach (FMCF) combines the metric distance computed by multiple operators from different channels of HSV color space with the factor (α, β, γ) as weight. In this experiment, we report the performance of FMCF with different factors of weight in Table 4(1) and Table 4(2). It is shown that the matching rate with $\alpha=0.5, \beta=0.3, \gamma=0.2$ is higher than other factors of weight at rank = 1, 5, 10, achieving 42.7%,71.3% and 68.1%. Among them, we can see that the weight of SWH-H is obviously higher than other operators, because it fuses the information of H–channel and S–channel. In contrast, the difference between of LDO-V and LEP-V is small.

Table 4(1)
The rank 1, 5, 10 matching rate (%) with our approach (FMCF) with different factors of weight on i-LIDS MCTS dataset.

α	0.1											
β	0.1				0.3				0.5			
γ	0.1	0.2	0.3	0.5	0.1	0.2	0.3	0.5	0.1	0.2	0.3	0.5
Rank-1	36	31	31.2	25.9	35	34.9	32.9	30.1	32.1	34.7	33.6	29.1
Rank-5	64.4	62.7	56.9	51.6	65.1	63	62.4	56	60	61.6	61	57.6
Rank-10	77.6	77.3	73.7	71.1	78.7	76.9	77.4	74.3	77	77	77.9	75.6

Table 4(2)
The rank 1, 5, 10 matching rate (%) with our approach (FMCF) with different factors of weight on i-LIDS MCTS dataset.

α	0.5											
β	0.1				0.3				0.5			
γ	0.1	0.2	0.3	0.5	0.1	**0.2**	0.3	0.5	0.1	0.2	0.3	0.5
Rank-1	35.1	37.9	35.4	32.9	36.9	**42.7**	37.7	35	39.3	39.9	40.4	36
Rank-5	66.4	69.8	65.3	63.2	68.4	**71.3**	65.3	65.1	67.7	67	68.1	64.4
Rank-10	76.9	78.7	75.1	76.7	80.2	**83.6**	79	78.7	79.7	78.1	80.9	77.6

Fig. 8. Examples of person re-identification on CUHK-01 dataset.

4.2. Performances on CUHK-01 dataset

The CUHK-01 dataset contains 971 persons and 3884 images, each of whom has two images in each camera view. Camera A captures the frontal or back views of pedestrians, whereas camera B captures their side views. All images are normalized to 60×160 pixels. Note that we choose all images of $p = 334, 486, 971$ persons and randomly select only one image form each person for the gallery set. In this dataset, the main problem of person re-identification is the large scale changes in camera view, shown in Fig. 8.

Also, we compare the rank-1, 5, 10, 20 matching rate of our proposed approach with several state-of-the-art and correlative approaches (color histogram, HOG, LBP, ELF, LEP and PRICoLBP), and report the results in Fig. 9 and Table 5. We can see that the matching rates at rank-1 with $p = 334, 486, 971$ are 26.5%, 25.0%, 22.1% which outperform the best result of 22.2%, 21.1%, 18.8% obtained by other approaches.From Fig. 9, it is noted that our proposed approach is significantly better than other approaches. It proves that our proposed approach(FMCF) can effectively extract the reliable features of color, texture and spatial information from all of the cell-regions in the person images which improve the adaptability of the large scale changes of view angle. Therefore, our proposed approach (FMCF) is most effective for the CUHK-01 dataset. Meanwhile, the SD/RR curve also has a better performance for the probability that any of the m-best matches is correct than other approaches.

5. Conclusion and future work

In this paper, we have presented an efficient and effective method for person re-identification. We have proposed a reliable and robust

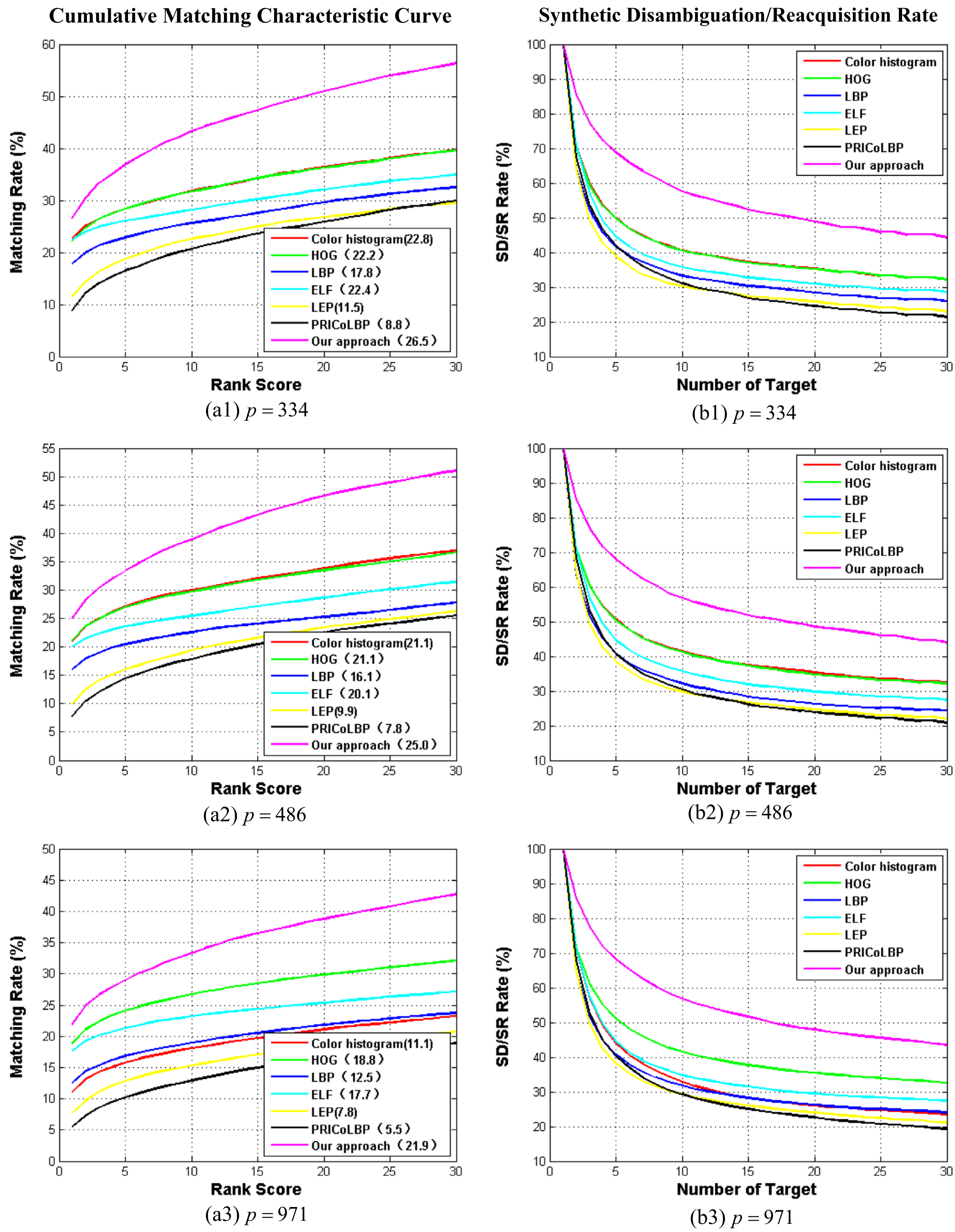

Fig. 9. The CMC and SD/RR curves of our approach, color histogram, LBP, HOG, ELF, LEP and PRICoLBP on CUHK-01 dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

descriptor called FMCF, which is shown to be effective for the viewpoint changes and illumination variations. The descriptor of FMCF pays more attention to the characteristics of different channels with HSV color space, combining color, texture and spatial information. Besides, we have discussed the performance, undergoing different scales of cell for capturing the 1-D feature histogram. Experiment on two challenging person re-identification datasets, i-LIDS MCTS and CUHK-01, show that the proposed method (FMCF) improves the rank-1 identification rates of correlative methods by 10.3% and 3.7% on the two databases, respectively. Due to the promising performance of the FMCF feature, it would be interesting to study other local features (e.g. Gabor, other color descriptors, etc.) or feature coding approaches with the same FMCF idea for person re-identification. It is also interesting to see the application of FMCF to other cross-view matching problems, such as the heterogeneous face recognition. In addition, we can bring metric learning idea into our proposed method for a better improvement of performance for person re-identification.

Table 5
The rank 1, 5, 10, 20 matching rate (%) with our approach, color histogram, LBP, HOG, ELF, LEP and PRICoLBP on CUHK-01 dataset.

Methods	$p = 334$				$p = 486$				$p = 971$			
	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 1$	$r = 5$	$r = 10$	$r = 20$
Our	**26.5**	**36.9**	**43.3**	**51.0**	**25.0**	**33.4**	**39.0**	**46.6**	**21.9**	**29.0**	**33.3**	**38.8**
Color histogram	22.8	28.4	31.8	36.4	21.1	27.2	29.6	33.8	11.1	15.1	18.1	21.1
HOG	22.2	28.3	31.7	36.3	21.1	26.9	29.7	33.5	18.8	24.1	26.7	29.9
LBP	17.8	22.9	25.7	29.6	16.1	20.5	22.6	25.3	12.5	16.8	18.9	21.8
ELF	22.4	26.1	28.2	32.1	20.1	23.6	25.5	28.6	17.7	21.3	23.2	25.3
LEP	11.5	18.7	22.6	26.8	9.9	16.0	19.4	23.7	7.8	13.0	16.0	18.7
PRICoLBP	8.8	16.5	20.7	26.0	7.8	14.4	17.8	22.5	5.5	10.2	12.9	16.5

Acknowledgements

The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions. This work is partially supported by China National Natural Science Foundation under Grant Nos. 61203247, 61273304, 61203376, 61202170, 61202318, 61573259 and 61472166. It is also supported by the Fundamental Research Funds for the Central Universities (Grant No. 2013KJ010). It is also partially supported by Changzhou Key Laboratory of Cloud Computing and Intelligent Information Processing grant No.CM20123004-KF01 and by the Open Project Program of Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education under Grant No. 30920130122005. It is also partially supported by the program of Further Accelerating the Development of Chinese Medicine Three Year Action of Shanghai Grant No. ZY3-CCCX-3-6002.

References

[1] A. Bedagkar-Gala, S.K. Shah, A survey of approaches and trends in person re-identification, Image Vision. Comput. 32 (4) (2014) 270–286.
[2] G. Doretto, T. Sebastian, P. Tu, et al., Appearance-based person reidentification in camera networks: problem overview and current approaches, J. Ambient Intell. Humaniz. Comput. 2 (2) (2011) 127–151.
[3] Du Y., Ai H., Lao S. Evaluation of color spaces for person re-identification, in: 21st International Conference on Pattern Recognition, 2012, pp. 1371-1374.
[4] Z.J. Xiang, Q. Chen, Y. Liu, Person re-identification by fuzzy space color histogram, Multimed. Tools Appl. 73 (1) (2014) 91–107.
[5] K. Jang, S. Han, I. Kim. Person Re-identification Based on Color Histogram and Spatial Configuration of Dominant Color Regions. arXiv: 1411.3410, (2014).
[6] Y. Li. Gabor-LBP based region covariance descriptor for person re-identification, in: Proceedings of the 2011 Sixth International Conference on Image and Graphics, 2011, pp.368-371.
[7] B. Jurie, Covariance descriptor based on bio-inspired features for person re-identification and face verification, Image Vision. Comput. 32 (6) (2014) 379–390.
[8] M. Verma, B. Raman, S. Murala, Local extrema co-occurrence pattern for color and texture image retrieval, Neurocomputing 165 (2015) 255–269.
[9] M. Farenzena, L. Bazzani, A. Perina, et al. Person re-identification by symmetry-driven accumulation of local features, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp.2360-2367.
[10] R. Zhao, W. Ouyang, X. Wang. Learning mid-level filters for person re-identification, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp.144-151.
[11] Y. Geng, H.M. Hu, G. Zeng, et al., A person re-identification algorithm by exploiting region-based feature salience, J. Vis. Commun. Image Represent. 29 (2015) 89–102.
[12] L. Bazzani, M. Cristani, V. Murino, Symmetry-driven accumulation of local features for human characterization and re-identification, Comput. Vision. Image Underst. 117 (2) (2013) 130–144.
[13] M. Zeng, Z. Wu, C. Tian, et al. Efficient Person Re-Identification by Hybrid Spatiogram and Covariance Descriptor, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015, pp.48-56.
[14] W.S. Zheng, S. Gong, T. Xiang, Reidentification by relative distance comparison, Pattern Anal. Mach. Intell. EEE Trans. 35 (3) (2013) 653–668.
[15] C.C. Guo, S.Z. Chen, J.H. Lai, et al. Multi-shot person re-identification with automatic ambiguity inference and removal, in: Proceedings of the 2014 22nd International Conference on Pattern Recognition, 2014, pp. 3540–3545.
[16] S. Iodice, A. Petrosino, Salient feature based graph matching for person re-identification, Pattern Recognit. 48 (4) (2015) 1070–1081.
[17] R. Zhao, W. Ouyang, X. Wang. Unsupervised salience learning for person re-identification, in: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3586–3593.
[18] S. Liao, Y. Hu, X. Zhu, et al. Person Re-identification by Local Maximal Occurrence Representation and Metric Learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2197–2206.
[19] S. Ding, L. Lin, G. Wang, et al., Deep feature learning with relative distance comparison for person re-identification, Pattern Recognit. 48 (10) (2015) 2993–3003.
[20] R. Zhao, W. Ouyang, H. Li, et al. Saliency detection by multi-context deep learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp.1265–1274.
[21] S. Z. Chen, C. C. Guo, J. H. Lai. Deep Ranking for Person Re-identification via Joint Representation Learning. arXiv preprint arXiv:1505.06821, 2015.
[22] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: Computer Vision–ECCV 2008, Springer, Berlin Heidelberg 2008, pp. 262–275.
[23] T. Lu, W. Shengjin, Person Re-Identification as Image Retrieval Using Bag of Ensemble Colors, IEICE Trans. Inf. Syst. 98 (1) (2015) 180–188.
[24] X. Wang, G. Doretto, T. Sebastian, et al. Shape and appearance context modeling, in: Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, 2007, pp.1-8.
[25] W. Hu, M. Hu, X. Zhou, et al., Principal axis-based correspondence between multiple cameras for person tracking, IEEE Trans. Pattern Anal. Mach. Intell. 28 (4) (2006) 663–671.
[26] B. Prosser, W.S. Zheng, S. Gong, et al., Person re-identification by support vector ranking, BMVC 2 (5) (2010) 6.
[27] F. Xiong, M. Gou, O. Camps, et al., Person re-identification using kernel-based metric learning methods, in: Computer Vision–ECCV 2014, Springer International Publishing 2014, pp. 1–16.
[28] S. Paisitkriangkrai, C. Shen, A. Hengel. Learning to rank in person re-identification with metric ensembles. arXiv preprint arXiv:1503.01543, 2015.
[29] P.M. Roth, M. Hirzer, M. Köstinger, et al., Mahalanobis distance learning for person re-identification, in: person re-identification, Springer, London 2014, pp. 247–267.
[30] D. Yi, Z. Lei, S. Liao, et al. Deep metric learning for person re-identification, in: Proceedings of the 2014 IEEE 22nd International Conference on Pattern Recognition, 2014, pp.34-39.
[31] G. Lisanti, I. Masi, A. Bagdanov, et al., Person re-identification by iterative re-weighted sparse ranking, IEEE Trans. Pattern Anal. Mach. Intell. 37 (8) (2015) 1629–1642.
[32] D. Gray,et al.. Evaluating appearance models for recognition, reacquisition, and tracking, in: Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS). 3(5) (2007).
[33] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[34] N. Dalal, B. Triggs. Histograms of oriented gradients for human detection, in: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp.886-893.
[35] A. Vadivel, S. Sural, A.K. Majumdar, An integrated color and intensity co-occurrence matrix, Pattern Recognit. Lett. 28 (8) (2007) 974–983.
[36] S. Murala, R.P. Maheshwari, R. Balasubramanian, Directional local extrema patterns: a new descriptor for content based image retrieval, Int. J. Multimed. Inf. Retr. 1 (3) (2012) 191–203.
[37] J. Qian, J. Yang, G. Gao, Discriminative histograms of local dominant orientation (D-HLDO) for biometric image feature extraction, Pattern Recognit. 46 (10) (2013) 2724–2739.
[38] G.H. Liu, J.Y. Yang, Content-based image retrieval using color difference histogram, Pattern Recognit. 46 (1) (2013) 188–198.
[39] X. Qi, R. Xiao, C.-G. Li, Y. Qiao, J. Guo, X. Tang, Pairwise rotation invariant co-occurrence local binary pattern, Pattern Analysis and Machine Intelligence, IEEE Trans. 36 (11) (2014) 2199–2213.
[40] M. Verma, B. Raman, S. Murala, Local extrema co-occurrence pattern for color and texture image retrieval, Neurocomputing 165 (2015) 255–269.
[41] F. Shen, C. Shen, Q. Shi, Anton van den Hengel, Z. Tang, H.T. Shen, Hashing on nonlinear manifolds, Image Process. IEEE Trans. 24 (6) (2015) 1839–1851.

[42] F. Shen, C. Shen, R. Hill, Anton van den Hengel, Z. Tang, Fast approximate $L\infty$minimization: Speeding up robust regression, Comput. Stat. Data Anal. 77 (2014) 25–37.

[43] F. Shen, C. Shen, Anton van den Hengel, Z. Tang, Approximate least trimmed sum of squares fitting and applications in image analysis, IEEE Trans. Image Process. 22 (5) (2013) 1836–1847.

[44] F. Shen, Z. Tang, J. Xu, Locality constrained representation based classification with spatial pyramid patches, Neurocomputing 101 (2013) 104–115.

[45] C. Zhao, D. Miao, Z. Lai, C. Gao, C. Liu, J. Yang, Two-dimensional color uncorrelated discriminant analysis for face recognition, Neurocomputing 113 (3) (2013) 251–261.

[46] Z. Lai, W.K. Wong, Y. Xu, C. Zhao, M. Sun, Sparse Alignment for Robust Tensor learning, IEEE Trans. Neural Netw. Learn. Syst. 25 (10) (2014) 1779–1792.

[47] C. Zhao, C. Liu, Z. Lai, Multi-scale gist feature manifold for building recognition, Neurocomputing 74 (17) (2011) 2929–2940.

[48] C. Zhao, Z.,C. Liu, X. Gu, J. Qian, Fuzzy local maximal marginal embedding for feature extraction, Soft Comput. 16 (1) (2012) 77–87.

[49] Z. Lai, W.K. Wong, Y. Xu, J. Yang, D. Zhang, Approximate orthogonal sparse embedding for dimensionality reduction, Trans. Neural Netw. Learn. Syst. 27 (4) (2016) 723–735.

[50] Z. Lai, Y. Xu, Z. Jin, D. Zhang, Human gait recognition via sparse discriminant projection learning, IEEE Trans. Circuits Syst. Video Technol. 24 (10) (2014) 1651–1662.

Xuekuan Wang is currently a master candidate in College of Electronics and Information Engineering, Tongji University. His research interests include computer vision, pattern recognition and machine learning, in particular, focusing on person re-identification for visual surveillance. E-mail:wxktongji@163.com.

Cairong Zhao is currently an associate professor at Tongji University. He received the Ph.D. degree from Nanjing University of Science and Technology, M.S. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, and B.S. degree from Jilin University, in 2011, 2006 and 2003, respectively. He is the author of more than 20 scientific papers in pattern recognition, computer vision and related areas. His research interests include computer vision, pattern recognition and visual surveillance. E-mail: zhaocairong@tongji.edu.cn.

Duoqian Miao is currently a full professor and vice dean of the school of Electronics and Information Engineering of Tongji University. He received his Ph.D. in Pattern Recognition and Intelligent System at Institute of Automation, Chinese Academy of Sciences in 1997. He works for Department of Computer Science and Technology of Tongji University, Computer and Information Technology Teaching Experiment Center, and the Key Laboratory of "Embedded System and Service Computing" Ministry of Education. He has published over 180 scientific articles in international journals, books, and conferences. He is committee member of International Rough Sets Society, senior member of China Computer Federation (CCF), committee member of CCF Artificial Intelligence and Pattern Recognition, committee member of Chinese Association for Artificial Intelligence (CAAI), chair of CAAI Rough Set and Soft Computing Society and committee member of CCAI Machine Learning, committee member of Chinese Association of Automation(CAA) Intelligent Automation, committee member and chair of Shanghai Computer Society(SCA) Computing Theory and Artificial Intelligence. His current research interests include: Rough Sets, Granular Computing, Principal Curve, Web Intelligence, and Data Mining etc. E-mail: dqmiao@tongji.edu.cn.

Zhihua Wei is currently an associate professor Tongji University. She received the double Ph.D. degrees from Tongji University, and Lyon University in 2010 simultaneously, M.S. degree and B.S. degree from Tongji University in 2005 and 2000. Her research interests include machine learning, image processing and data mining. E-mail: zhihua_wei@tongji.edu.cn.

Renxian Zhang is currently an associate professor at Tongji University. He received a Ph.D. in Computer Science from The Hong Kong Polytechnic University (2012) and a second one in Linguistics from Fudan University (2005). He is the author of more than 20 journal and conference papers in Natural Language Processing, Data Mining and related areas. His research interests include text cognitive modeling, automatic text summarization, and social media analysis. Email: rxzhang@tongji.edu.cn.